

基于多尺度 V1-MT 前馈模型的光流计算方法

张文达,许悦雷,马时平,李 帅,邹洪中

(空军工程大学 航空航天工程学院,西安 710038)

摘 要:针对在大位移、弱纹理等情况下光流信息计算误差较大的问题,提出一种基于改进 V1-MT 前馈模型的光流计算方法。以视频序列作为输入,分别计算初级视皮层和中颞叶(MT)细胞的响应,分解 MT 细胞的响应得到光流信息。采用多尺度和由粗到精的方法,解决大位移情况下的光流计算问题,采用基于双边滤波的插值方法,融合邻域运动信息,以估计弱纹理区的光流信息。仿真结果表明,与其他基于生物模型的计算方法相比,该方法能更为精确地计算光流信息。

关键词:光流;运动信息;大位移;多尺度;双边滤波

中文引用格式:张文达,许悦雷,马时平,等. 基于多尺度 V1-MT 前馈模型的光流计算方法[J]. 计算机工程,2017,43(9):205-209.

英文引用格式:ZHANG Wenda, XU Yuelei, MA Shiping, et al. Optical Flow Calculating Method Based on Multi-scale V1-MT Feedforward Model[J]. Computer Engineering, 2017, 43(9):205-209.

Optical Flow Calculating Method Based on Multi-scale V1-MT Feedforward Model

ZHANG Wenda, XU Yuelei, MA Shiping, LI Shuai, ZOU Hongzhong

(Institute of Aeronautics and Astronautics Engineering, Air Force Engineering University, Xi'an 710038, China)

[Abstract] In order to solve the problem of great computational error of optical flow information under large displacement and weak texture, an optical flow computational method based on an improved V1-MT feedforward model is proposed. By taking video sequences as inputs, the responses of cells in primary visual cortex and Middle Temporal(MT) are calculated. The optical flow information can be obtained by decomposing MT response. The multi-scale and coarse to fine methods are used to solve the optical flow computational problem under large displacement. The interpolation method based on bilateral filtering is adopted, so that neighborhood motion information can be fused and the optical flow information in the weak texture region can be estimated. Simulation results show that this method can calculate optical flow information more accurately than other methods based on biological model.

[Key words] optical flow; movement information; large displacement; multi-scale; bilateral filtering

DOI:10.3969/j.issn.1000-3428.2017.09.036

0 概述

随着美国和欧洲“脑计划”的兴起,针对哺乳动物大脑工作机制的探索与建模逐渐成为新的热点。如何将大脑中视觉处理的机制拓展运用到目标识别^[1]、运动信息提取^[2]以及动作识别^[3]等方面越来越受到人们的关注。

人类所获取的信息中超过 80% 都是通过视觉得到的,视觉是感知周围世界信息的重要渠道。灵长类哺乳动物的视觉通路有 2 条^[4]:“WHAT”通路,即腹侧通路,主要对颜色纹理形状等信息进行处理;“WHERE”通路,即背侧通路,主要对空间和运动信

息进行处理。

对周围环境中运动信息感知是哺乳动物视觉系统的重要功能。从上世纪 80 年代开始,学者们对视觉通路中运动信息的处理过程提出了一系列模型,主要是针对视觉皮层的初级视皮层 V1 区和中颞叶(Middle Temporal, MT)区。文献[5]提出了一种经典的级联模型,按照 V1 简单细胞、V1 复杂细胞和 MT 细胞分别进行建模,V1 简单细胞的感受野采用高斯函数的三阶导数,V1 复杂细胞的处理为半波整流与简单正则化,之后再作为 MT 细胞的输入。STA(Spike-triggered Average),STC(Spike-triggered Covariance)模型^[6]是一种数据驱动模型,通过求取

基金项目:国家自然科学基金(61372167,61379104)。

作者简介:张文达(1991—),男,硕士研究生,主研方向为模式识别、人工智能;许悦雷,教授;马时平,副教授;李 帅,博士研究生;邹洪中,硕士研究生。

收稿日期:2016-09-02

修回日期:2016-10-06

E-mail:1475337387@qq.com

产生触发脉冲的刺激的均值与方差来对细胞的滤波特性进行描述。iSTAC (Information-theoretic Spike-Triggered Average and Covariance) 则引入信息理论^[7], 通过最大化激励信号和细胞响应之间的互信息进行滤波器的选择。以上的工作主要针对生理实验所记录的观测数据提出一系列模型进行解释, 模型中所采用的输入一般为随机点、光栅或者光格, 不能直接应用于复杂的自然场景。

将视皮层中对运动信息的处理机制运用到复杂的自然场景中可以为计算机视觉提供新的研究思路, 是目前一个新的研究热点。文献[8]基于疏架构及递归机制, 提出了一种基于 V1-MT 递归模型的运动估计算法, 可以实现真实场景下的运动分割问题; 文献[9]从大脑视觉通路特性出发, 利用硬件实现了一种基于多通路梯度模型的鲁棒的运动估计算法; 文献[10]从无脊椎动物视觉通路得到启发, 提出了一种基于 firing-rate 机制的运动估计模型; 文献[11]基于 V1-MT 前馈模型, 提出了一种通过解码 MT 细

胞响应进行运动估计的方法。但是, 以上方法针对大位移和弱纹理区的处理能力不足, 难以对真实场景下的运动信息进行精确估计。

本文在文献[5]级联模型的基础上, 首先输入相应的图像序列, 得到 V1 和 MT 区细胞的响应; 然后对 MT 区细胞的响应进行分解得到各个位置的速度矢量-光流。为了更好地应对大位移情况下的光流估计问题, 引入多尺度和由粗到细的策略; 在弱纹理区采用一种基于双边滤波的插值方法, 弱纹理区运动信息通过融合其邻域运动信息进行估计。

1 V1-MT 前馈模型

本文模型在结构上分为 V1 和 MT 2 个阶段, 分别对应背侧通路的最重要的 2 个视皮层区域的处理流程, 在分别求解 V1 细胞和 MT 细胞响应之后, 通过分解 MT 细胞组的响应来得到光流信息, 如图 1 所示。

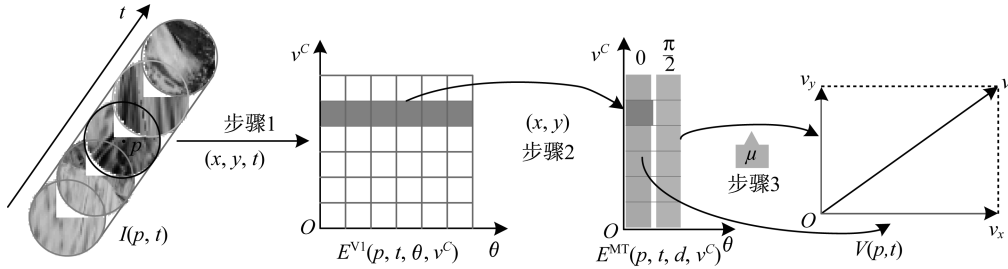


图 1 本文模型流程

1.1 V1 细胞响应

V1 细胞按照感受野的大小可以分为简单细胞和复杂细胞, 主要对视觉刺激的朝向和与之正交的运动速度进行调谐^[12]。

V1 简单细胞的感受野通常由时空域中的带通滤波器 $g(x, y, t; \theta, v^c, \sigma)$ 描述。目前已经提出了很多不同的时空滤波器模型。文献[5]采用高斯函数的三阶导数来描述时空滤波器, 文献[13]则采用 Gabor 函数的一阶和二阶导数来描述时空滤波器。为了降低计算的复杂度, 可以将之分解为时间域和空间域中不同的滤波器。其中空间域中的滤波器如式(1)所示。

$$h(x, y) = B e^{-(x^2 + y^2)/2\sigma^2} e^{j2\pi(f_{x0}x + f_{y0}y)} \quad (1)$$

其中, $f_{x0} = f_0 \cos \theta$; $f_{y0} = f_0 \sin \theta$; f_0 为空间频率; σ 为空间尺度参数。

时间域中的滤波器如式(2)所示。

$$p(t) = e^{(-t/\tau)} e^{j2\pi(f_{t0}t)} \quad (2)$$

其中, $f_{t0} = f_0 v^c$; τ 是时间尺度参数。

记复滤波器 h 和 p 的实部和虚部分别为 h_e, p_e 和 h_o, p_o , 则时空滤波器的实部和虚部分别为:

$$g_e(x, y, t) = h_e(x, y)p_e(t) - h_o(x, y)p_o(t) \quad (3)$$

$$g_o(x, y, t) = h_o(x, y)p_e(t) - h_e(x, y)p_o(t) \quad (4)$$

偏好方向为 θ 、偏好速度为 v^c 的 V1 简单细胞的响应为输入视频序列与对应滤波器的卷积, 即:

$$R_{e/o}(x, y, t; \theta, v^c, \sigma) = g_{e/o}(x, y, t; \theta, v^c, \sigma) * I(x, y, t) \quad (5)$$

V1 简单细胞的响应的能量函数为:

$$E(x, y, t; \theta, v^c, \sigma) = R_e^2(x, y, t; \theta, v^c, \sigma) + R_o^2(x, y, t; \theta, v^c, \sigma) \quad (6)$$

经过正则化后可得 V1 复杂细胞的响应:

$$E(x, y, t; \theta, v^c, \sigma) = \frac{E(x, y, t; \theta, v^c, \sigma)}{\sum_{\theta_i = \theta_1}^{\theta_N} E(x, y, t; \theta_i, v^c, \sigma) + \varepsilon} \quad (7)$$

其中, ε 是一个很小的正常数, 使得在没有运动能量的区域分母不为 0。

1.2 MT 细胞响应

MT 细胞接受 V1 细胞的输入, 其感受野比 V1 细胞的大。MT 细胞主要调谐运动的方向和速率大小, 而与在图像中相对的朝向无关。这种功能主要通过通过对 V1 输出分别在空间域和朝向域进行采样, 以

及非线性正则化来实现^[5]。

偏好方向为 d 、偏好速度为 v^c 的 MT 细胞的响应为:

$$R^{\text{MT}}(x, y, t; v^c, d, \sigma) = \sum_{\theta_i=\theta_1}^{\theta_N} w_d(\theta_i) G(\cdot)^{(x, y, t)} * \tilde{E}(\cdot) \quad (8)$$

其中, $G(\cdot)$ 为高斯核, 代表空间域的采样函数; $w_d(\theta) = \cos(d - \theta)$, $d \in [0, 2\pi]$, 代表空间朝向域的采样函数; $*$ 表示卷积。

MT 细胞的能量函数为:

$$E^{\text{MT}}(x, y, t; v^c, d, \sigma) = F(R^{\text{MT}}(x, y, t; v^c, d, \sigma)) \quad (9)$$

其中, $F(\cdot)$ 为 e 指数函数, 是一个非线性函数。

2 运动信息提取

2.1 运动速度估计

速度矢量并不能从唯一地从单个 MT 细胞的响应获得, 但是可以通过分解 MT 细胞组的响应来得到^[14]。在各个方向上的运动分量可由式(10)得到:

$$v^d(x, y, t; d, \sigma) = \sum_{v_i^c=v_1^c}^{v_M^c} v_i^c E^{\text{MT}}(x, y, t; v_i^c, d, \sigma) \quad (10)$$

根据交叉约束 (Intersection of Constraints, IOC) 准则, 可定义如下损失函数^[15]:

$$G(\mathbf{v}) = \sum_{d_i=d_1}^{d_N} (v^d(x, y, t; d_i, \sigma) - v[\cos d_i \sin d_i])^2 \quad (11)$$

则上式中运动矢量 \mathbf{v} 为:

$$\mathbf{v} = \operatorname{argmin} \{ G(\mathbf{v}) \} \quad (12)$$

其最小二乘解就是求解如下方程组:

$$\begin{cases} \frac{\partial G(\mathbf{v})}{\partial x} = 0 \\ \frac{\partial G(\mathbf{v})}{\partial y} = 0 \end{cases} \quad (13)$$

分别求损失函数对 v_x 和 v_y 的偏导数:

$$\frac{\partial G(\mathbf{v})}{\partial x} = -2 \sum_{d_i=d_1}^{d_N} (v^d(x, y, t; d_i, \sigma) - v_x \cos d_i - v_y \sin d_i) \cos d_i \quad (14)$$

$$\frac{\partial G(\mathbf{v})}{\partial y} = -2 \sum_{d_i=d_1}^{d_N} (v^d(x, y, t; d_i, \sigma) - v_x \cos d_i - v_y \sin d_i) \sin d_i \quad (15)$$

N 表示在运动方向空间 $[0-\pi]$ 均匀采样的点数, 由于:

$$\sum_{d_i=d_1}^{d_N} \sin^2 d_i = \sum_{d_i=d_1}^{d_N} \cos^2 d_i = N/2 \quad (16)$$

$$\sum_{d_i=d_1}^{d_N} \sin d_i \cos d_i = 0 \quad (17)$$

因此式(13)的解为:

$$v_x = \frac{2}{N} \sum_{d_i=d_1}^{d_N} v^d(x, y, t; d_i, \sigma) \cos d_i \quad (18)$$

$$v_y = \frac{2}{N} \sum_{d_i=d_1}^{d_N} v^d(x, y, t; d_i, \sigma) \sin d_i \quad (19)$$

为了简化计算, 取 $N=2$, 即 $d_1=0$, $d_2=\pi/2$, 代入式(18)、式(19), 可以得到:

$$v_x = \sum_{v_i^c=v_1^c}^{v_M^c} v_i^c E^{\text{MT}}(x, y, t; v_i^c, 0, \sigma) \quad (20)$$

$$v_y = \sum_{v_i^c=v_1^c}^{v_M^c} v_i^c E^{\text{MT}}(x, y, t; v_i^c, \pi/2, \sigma) \quad (21)$$

2.2 基于多尺度的光流估计

第 1 节中的 V1-MT 前馈模型最初是用来解释生理实验中所得到的神经元的响应, 所用的激励信号一般为光栅或者光格, 为了减少模型的计算量, 只采用了单一的空间频率 f_s 。而自然图像中的信息是分布在一个很宽的频率范围; 同时, 在自然场景中往往存在大位移的情况。这就需要有一种机制来从整个频率范围获取信息, 应对自然场景中存在的大位移问题。

多尺度是计算机视觉中比经典的一种方法, 包括金字塔分解和由粗到细的调整过程^[16], 如图 2 所示。

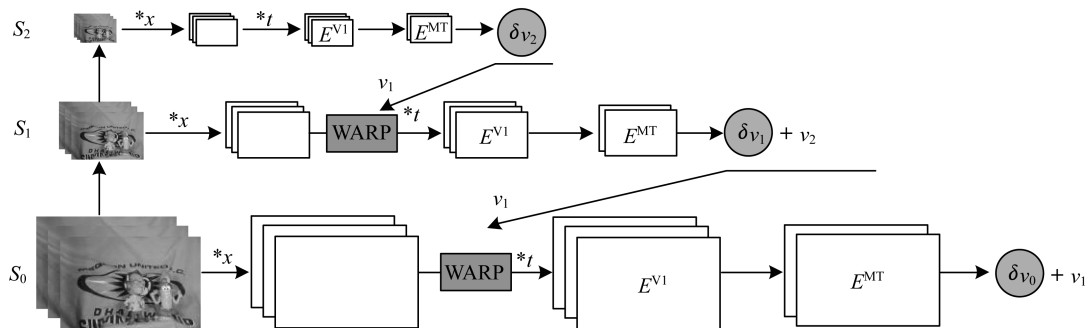


图 2 多尺度的光流估计

首先将每一帧图像进行金字塔分解得到各层图像 $S_l (l=0, 1, \dots, L-1)$, 其中, $l=0$ 为原始图像。在金字塔分辨率较低的图像层计算得到 δv_l , 扩展后得到 v_l , 并与 $l-1$ 层经过空间滤波之后的图像求得 $l-1$ 层的残存光流 (Residual Optic Flow, ROF), 可以

得到 $l-1$ 层的光流信息为:

$$v_{l-1} = \delta v_{l-1} + v_l \quad (22)$$

经过这种多尺度的方法, 在分辨率较低的图像层中, 空间距离被拉近了, 可以更好地应对大位移情况下的光流估计问题。

2.3 边界区域及无纹理区域的光流估计

如图 3 所示,图 3(a)中 p_1 表示 V1 细胞的感受野, p_2 表示 MT 细胞的感受野, \hat{A}_1, \hat{A}_2 部分代表边界区域;图 3(b)中分别表示需要插值估计的区域—边界区域和无纹理区;图 3(c)为边界区域的插值示意图;图 3(d)为弱纹理区的插值示意图。其中,区域 A 是 Ω_{in} 中临近边界或者无纹理区若干像素的一个区域。对于弱纹理区域或者边界区域等区域的运动信息,可以用其邻域的运动信息来估计。

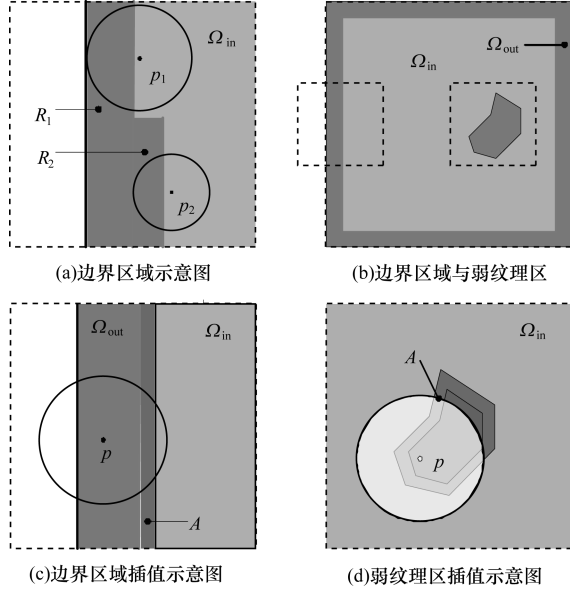


图 3 边界及弱纹理区插值示意图

对于边界区域和弱纹理区:

$$E^{MT}(p) = \frac{1}{N(p)} \int_{p' \in A} f_{\alpha}(\|p - p'\|) f_{\gamma}(I(p', t) - I(p, t)) dp' \quad (23)$$

其中:

$$N(p) = \int_{p' \in A} f_{\alpha}(\|p - p'\|) f_{\gamma}(I(p', t) - I(p, t)) dp' \text{ 是一个正则项。}$$

3 仿真实验与结果分析

3.1 仿真数据库

为了验证所提出方法的性能,选用目前比较流行的 Middlebury 光流测试库。该测试库提供了涵盖实验室真实场景、合成场景、高速真实场景、立体摄影等 24 个图像序列,规格上有灰度/彩色,2 帧/8 帧之别。其中, Mequon, Schefflera, Wooden, Urban, Yosemite, Teddy 等序列的 Ground-truth 数据隐藏,

RubberWhale, Hydrangea, Dimetrodon, Grove2, Grove3, Urban2, Urban3, Venus 等序列的 Ground-truth 开放,便于研究者进行算法性能对比。

3.2 参数设置

本文中 V1 细胞感受野的空间尺度参数 σ 设为 2.5 像素,时间尺度参数 τ 设为 2.5 帧,空间频率 f_0 为 2.5 像素/周期^[17]。时间频率 f_t 为 $\{0, 0.10, 0.15, 0.23\}$ 圈/帧, V1 细胞调谐方向个数 N 为 8, V1 细胞的调谐方向 θ_i 为 $k\pi/N, k = 0, 1, \dots, N-1$, 常数 ε 为 10^{-9} , MT 细胞调谐速度个数 M 为 7, 调谐速度 v^c 为 $\{-0.9, -0.6, -0.4, 0, 0.4, 0.6, 0.9\}$ 。金字塔分解的层数 L 取为 5。

3.3 实验结果分析

光流精度是衡量光流畅质量的指标,一般可以用计算得到的光流值和图像运动值(即光流场的真实值之间的)误差表示。其评价指标主要有角误差(Angular Error, AE)和终点误差(Endpoint Error, EPE)等。

角误差是 2 个光流 (u_x, u_y) 和 (u_{xG}, u_{yG}) 对应的 3D 向量 $OP = (u_x, u_y, 1)^T$ 和 $OP_G = (u_{xG}, u_{yG}, 1)^T$ 之间的角,通常用归一化的向量计算点积,然后再经过反余弦运算获得:

$$AE = \arccos\left(\frac{OP_G^T \cdot OP}{\|OP_G\| \cdot \|OP\|}\right) \quad (24)$$

其中, (u_x, u_y) 是算法估计出的光流场; (u_{xG}, u_{yG}) 是光流场的真实值(Ground Truth)。角误差的单位可以是弧度也可以是角度,其平均值可以衡量整场的平均值。

终点误差是绝对值,定义为估计光流场与真实光流场之间距离的偏差来作为判断标准:

$$EPE = \sqrt{(u_x - u_{xG})^2 + (u_y - u_{yG})^2} \quad (25)$$

本文方法选取平均角误差(Average Angular Error, AAE)和平均终点误差(Average Endpoint Error, AEE)作为质量评价指标。

本文选取文献[8, 11]的算法作为对比实验,结果如表 1 所示。从表 1 中可以看出,本文方法在平均角误差和平均端点误差这 2 个参数上均小于对比算法,在光流计算精度上有较大幅度的提高。而本文方法在 RubberWhale、Army 和 Urban23 个序列的误差相对比较大,是由于这 3 个序列中细节信息丰富,尖锐的边缘处较多。

表 1 文献[8, 11]及本文方法光流计算性能对比

视频序列	文献[8]算法		文献[11]算法		本文方法	
	AAE/(°)	AEE/(像素·s ⁻¹)	AAE/(°)	AEE/(像素·s ⁻¹)	AAE/(°)	AEE/(像素·s ⁻¹)
RubberWhale	15.35	0.52	10.20	0.34	9.65	0.29
Army	16.38	0.61	12.02	0.33	11.54	0.28
Grove	9.27	1.05	6.51	0.76	5.97	0.66
Hydrangea	7.94	0.95	5.96	0.62	5.89	0.57
Urban2	18.02	1.73	14.51	1.46	13.96	1.38

本文方法的仿真结果如图4所示。第1列分别为各序列的第一帧,第2列为对应的光流场的真实值,第3列为本文方法所估计得到的光流场,第4列为本文方法所对应的角误差的分布图,其中越亮的位置代表其角误差越大。

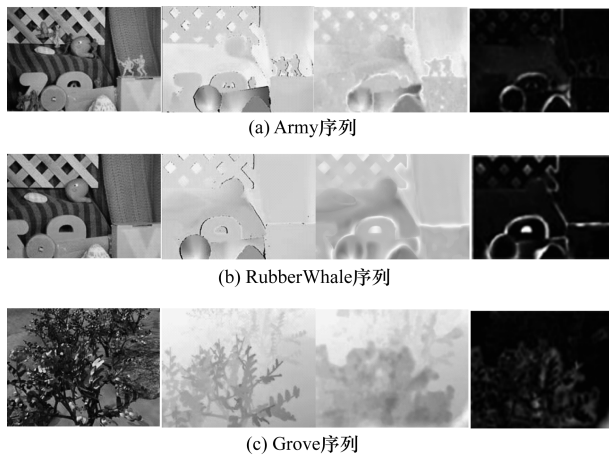


图4 仿真结果

从图4中第3列和第4列可以看出,本文方法对序列图片边缘处、弱纹理区域都可以较好地对运动信息进行估计,而误差主要集中在纹理信息丰富区域以及尖锐的边缘区域。本文方法采用多尺度地由粗到细的策略,可以很好地应对大位移问题,使得在位移较大处的运动估计误差没有明显增大。而在序列图像边界区域以及弱纹理区域采用了基于双边滤波的插值方法,通过融合其邻域内的运动信息来估计序列图像边界区域以及弱纹理区域的运动信息,使得序列图像的边界区域和弱纹理区域的运动信息得到较为准确的估计。而由于在 V1-MT 前馈模型中 V1 简单细胞的感受野是基于 Gabor 滤波器的,其响应是由序列图像与 Gabor 滤波器进行卷积得到的。在计算简单细胞的响应时较大的感受野可以更好地感知整体轮廓运动信息而忽略细节信息,较小的感受野可以更好地感知细节信息而容易忽略整体轮廓运动信息。而本文采用固定大小的感受野,是导致纹理信息丰富区域以及尖锐边缘区域误差较大的主要原因。

4 结束语

本文将经典的 V1-MT 前馈模型应用到视频序列运动信息的提取,提出一种光流计算方法。根据 V1 和 MT 细胞的级联特性。分别求解 V1 和 MT 区的细胞响应,通过分解 MT 响应得到光流信息。实验结果表明,该方法能够有效提取运动信息。下一步将研究如何自适应地调节感受野的范围来更加精确地进行运动估计。

参考文献

- [1] 李岳云,许悦雷,马时平,等. 仿人脑视皮层机制的目标识别方法[J]. 计算机工程与设计,2015,36(8): 2147-2152.
- [2] BART K, RICHARD J A. Neural Mechanisms of Speed Perception: Transparent Motion [J]. Journal of Neurophysiology, 2013, 110(9): 2007-2018.
- [3] 刘雨娇,范勇,高琳,等. 基于时空深度特征的人体行为识别算法[J]. 计算机工程,2015,41(5): 259-263.
- [4] UNGERLEIDER L G, MISHKIN M. Two Cortical Visual Systems[J]. Analysis of Visual Behavior, 1982, 8: 549-586.
- [5] SIMONCELLI E P, HEEGER D J. A Model of Neuronal Responses in Visual Area MT [J]. Vision Research, 1998, 38(5): 743-761.
- [6] SCHWARTZ O, PILLOW J W, RUST N C, et al. Spike-triggered Neural Characterization[J]. Journal of Vision, 2006, 6(4): 484-507.
- [7] PILLOW J W, SIMONCELLI E P. Dimensionality Reduction in Neural Models: An Information-theoretic Generalization of Spike-triggered Average and Covariance Analysis[J]. Journal of Vision, 2006, 6(4): 414-428.
- [8] BAYERL P, KNUTSSON H. A Fast Biological Inspired Algorithm for Recurrent Motion Estimation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(2): 246-260.
- [9] BOTELLA G, GARCIA A. Robust Bioinspired Architecture for Optical Flow Computation [J]. IEEE Transactions on Very Large Scale Integration Systems, 2010, 18(4): 616-629.
- [10] DELLEN B, WESSEL R, CLARK J. Motion Processing with Wide-field Neurons in the Retino-tecto-rotundal Pathway[J]. Journal of Computing Neuroscience, 2010, 28(1): 47-64.
- [11] SOLARI F, CHESSA M, MEDATHATI N V K, et al. What Can We Expect from a V1-MT Feedforward Architecture for Optical Flow Estimation [J]. Signal Processing Image Communication, 2015, 39(6): 342-354.
- [12] ADELSON E H, MOVSHON J A. Phenomenal Coherence of Moving Visual Patterns[J]. Nature, 1982, 300(5892): 523-525.
- [13] ADELSON E H, BERGEN J. Spatiotemporal Energy Models for the Perception of Motion[J]. Journal of the Optical Society of America A, 1985, 2(2): 284-299.
- [14] POUGET A, ZHANG K, DENEVE S, et al. Statistically Efficient Estimation Using Population Coding [J]. Neural Computation, 1998, 10(2): 373-401.
- [15] RAD K R, PANINSKI L. Information Rates and Optimal Decoding in Large Neural Populations[C]//Proceedings of Advances in Neural Information Processing Systems. Granada, Spain: IEEE Press, 2011: 846-854.
- [16] SIMONCELLI E P. Course-to-fine Estimation of Visual Motion[C]//Proceedings of the 8th IEEE Workshop on Image and Multidimensional Signal Processing. Washington D. C., USA: IEEE Press, 1993.
- [17] CHESSA M, SABATINI S P, SOLARI F. A Systematic Analysis of a V1 - MT Neural Model for Motion Estimation[J]. Neurocomputing, 2016, 173(P3): 1811-1823.

编辑 刘冰